**MAJOR PROJECT REPORT**

**on**

**DENGUE CASES PREDICTION WITH ENVIRONMENTAL FACTORS USING MACHINE LEARNING**



**Submitted in**

**partial fulfilment of the requirements**

**for the award of**

**Bachelor of Technology**

**in**

**Information Technology**

**BY**

K. Hemanth Kumar [B18IT032]

A. Harsha Nandan [B18IT013]

D. Ragini [B18IT060]

K. Soumya [B18IT032]

**Under the Supervision of**

Sri. P. Sudharshan Ray

Associate Professor

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**KAKATIYA INSTITUTE OF TECHNOLOGY & SCIENCE, WARANGAL-15**

**2021 - 2022**



**C E R T I F I C A T E**

This is to certify that **K. Hemanth Kumar (B18IT029), A. Harsha Nandan (B18IT013), D. Ragini (B18IT060), K. Soumya (B18IT029)** of VII-Semester B. Tech Information Technology has satisfactorily completed the **Major Project Phase-1 (U18IT707)** entitled **“DENGUE CASES PEDICTION WITH ENVIRONMENTAL FACTORS USING MACHINE LEARNING”** in the partial fulfilment of the requirement of B. Tech degree during this academic year 2021-2022.

|  |  |
| --- | --- |
| **SUPERVISOR** | **CHAIRMAN, DMPEC** |
| Sri. P. SUDHARSHAN RAY  **Associate Professor, Dept of I.T.** | Dr. P. KAMAKSHI  **Professor & Head, Dept. of I.T.** |

**ACKNOWLEDGEMENT**

We wish to take this opportunity to express our deep gratitude to all the people who have extended their cooperation in various ways during our Major Project Phase-1 (U18IT707). It is our pleasure to acknowledge the help of all those individuals.

We thank **Dr. K. Ashoka Reddy**, Principle of Kakatiya Institute of Technology & Science, Warangal, for his strong support.

We thank **Dr. P. Kamakshi**, Professor & Head, Department of Information Technology for her constant support in bringing shape of this Major Project Phase-1 (U18IT707).

We would like to thank our supervisor, **Sri. P. Sudharshan Ray**, Associate Professor of Information Technology for his guidance and help throughout the Major Project Phase-1 (U18IT707).

In completing this Major Project Phase-1 (U18IT707) successfully all our faculty members have given an excellent cooperation by guiding us in every aspect. All your guidance helped us a lot and we are very grateful to you.

K. SOUMYA [B18IT032]

A. HARSHA NANDAN [B18IT013]

D. RAGINI [B18IT060]

K. HEMANTH KUMAR [B18IT029]

**ABSTRACT**

Dengue fever is one of the well popular mosquito-borne sicknesses that occurs in tropical and subtropical bits of the world. The transmission of dengue can be related to climatic elements since it is spread by mosquitoes. Using the normal data accumulated by various Government associations we endeavour to expect the amount of dengue fever cases declared each week in two metropolitan networks San Juan, Puerto Rico and Iquitos, Peru. This venture expects to configuration time series based ARIMA model utilizing various boundaries like temperature, vegetation and precipitation information and consolidating time series, aspect decrease for better expectation of dengue flare-up. Exact dengue expectations would help general wellbeing laborers and individuals all throughout the planet find ways to lessen the effect of these pandemics. However, anticipating dengue is a strong undertaking that requires the union of various informational indexes on sickness occurrence, climate, and the climate, a comprehension of the connection among environment and dengue elements can further develop research drives and asset designation to assist with battling hazardous pandemics.

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Titles** | **Page No.** |
| **1** | **INTRODUCTION** | **1** |
|  | 1.1 Introduction | 1 |
| **2** | **LITERATURE SURVEY** | **3** |
|  | 2.1 Existing System | 3 |
|  | 2.2 Proposed System | 4 |
| **3** | **DESIGN** | **5** |
|  | 3.1 System Requirements Specifications  3.1.1 Software Requirement Specifications  3.1.2 Hardware Requirement Specifications | 5  5  5 |
|  | 3.2 Time Series Modelling | 5 |
|  | 3.3 UML Diagrams  3.3.1 Activity diagram for ARIMA model  3.3.2 Activity diagram for LSTM model | 6  6 |
| **4** | **MODULES DESCRIPTION** | **7** |
|  | 4.1 Importing Data  4.1.1 Describing the Data | 7 |
|  | 4.2 Data Pre-Processing  4.2.1 Checking Null values  4.3 Exploratory Data Analysis  4.3.1 Seasonal Decompose | 7  9 |
|  | 4.4 Stationary Test  4.4.1 Visualization  4.4.2 Dickey-Fuller Test  4.5 ARIMA Implementation  4.5.1 ACF & PACF plots  4.5.2 Model building  4.5.3 Model evaluation  4.6 LSTM Implementation  4.6.1 Importing data  4.6.2 Heat map  4.6.3 Forecast for San Juan data  4.6.4 Forecast for Iquitos data | 12 |
| **5** | **RESULTS** | **15** |
| **6** | **CONCLUSION AND FUTURE SCOPE** | **17** |
| **7** | **REFERENCES** | **18** |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Fig No.** | **Figure Name** | **Page No.** |
| 3.1 | Activity Diagram for ARIMA Model | 6 |
| 3.2 | Activity Diagram for LSTM Model | 7 |
| 4.1 | Importing Data | 8 |
| 4.2 | Describing Data | 8 |
| 4.3 | Checking Null Values | 9 |
| 4.4 | Filling Null Values | 10 |
| 4.5 | Cases in San Juan | 11 |
| 4.6 | Seasonal Decomposition Graph | 12 |
| 5.1 | Rolling Mean & Standard Deviation | 16 |
| 5.2 | Dickey Fuller Test Result | 17 |

**INTRODUCTION**

**1.1 INTRODUCTION**

Dengue fever is a mosquito-borne tropical illness transmitted by the dengue virus. Symptoms usually appear three to fourteen days after the illness. Excessive fever, headache, vomiting, muscle and joint pains, and a brand-name skin rash are all possible symptoms. The average recovery time is seven days. In a small percentage of cases, the disease progresses to a more serious dengue hemorrhagic fever, with biting the dirt, low levels of blood platelets, and blood plasma spillage, or to dengue shock, with dangerously low circulatory strain. The most precise depictions of an episode date back to 1779. Its viral origins and spread were first noticed in the mid-twentieth century. Dengue fever has been a global problem since the Second World War, and it is now found in over a hundred and twenty countries. for the most part in Southeast Asia, South Asia and South America. Around 390 million individuals are tainted a year, about a large portion of 1,000,000 require emergency clinic affirmation, and roughly 40,000 pass-on. In 2019 a huge expansion in the quantity of cases was seen. Aside from taking out the mosquitos, work is progressing for prescription designated straightforwardly at the infection. It is named a disregarded tropical infection.

Surrounding temperature and precipitation are likewise significant elements that straightforwardly influence the advancement of dengue infection in significant Environmental change won't just influence the pace of mosquito improvement, yet in addition the infection brooding time. At the end of the day, climatic elements impact dengue nature both straightforwardly and by implication by influencing mosquito development elements, infection replication, and mosquito–human collaborations

Some new assessments have moreover recollected spatial part for showing climate fever conjecture reliant upon Bayesian most noteworthy entropy examination to survey the climatic ramifications for dengue appointment in southern non mainland china. Applying a spacetime subject to the data for the 2002 to 2006 period, they found basic positive associations between is precipitation, least temperature, and dengue recurrence. The expected spatiotemporal dengue fever scattering was furthermore especially close to the genuine allotment of dengue cases uncovered for the year 2007.

Notwithstanding environment conditions, populace thickness and urbanization are likewise viewed as significant driving elements for. For instance,. demonstrated that populace thickness fundamentally affects the quantity of detailed provinces in Taiwan. Furthermore, guaranteed that fast populace development in tropical metropolitan regions regularly gives ideal biological conditions to Ae. aegypti numbers to increment. By and large, unique logical methodologies are applied relying upon the distributional suspicions (e.g., Poisson, ordinary) and the spatial as well as fleeting elements of the reaction.

**CHAPTER 2**

**LITERATURE SURVEY**

It is one of the significant assignments to demonstrate dengue fever cases to help general wellbeing officials to design and set up their money chests to help dengue fever episode. displaying of aggregated procured at the location called San Juan. Assessment of the cast for unborn episode outcomes, as validation is introduced for the pattern and occasional flare-ups in San Juan.

* 1. **EXISTING SYSTEM**

Creating robust and effective forecasting models is troublesome due to the intricacies of mosquito nature and infection transmission cycles. They tended to this test with an original methodology that utilizes a transformative calculation to streamline a spatiotemporal jungle fever model by apportioning noticed areas into clusters, every one of which addresses an alternate transmission climate with an interesting arrangement of natural sensitivities caught by the model. Time series models of intestinal sickness cases can be applied to gauge scourges and backing proactive mediations. Mosquito life history and parasite advancement are touchy to natural factors like temperature and precipitation, and these factors are regularly utilized as indicators in jungle fever models. Be that as it may, jungle fever climate connections can shift with environmental and social setting. They utilized a hereditary calculation to upgrade a spatiotemporal intestinal sickness model by conglomerating areas into clusters with comparative natural sensitivities. They tried the calculation in the Amhara Region of Ethiopia utilizing seven years of week after week Plasmodium falciparum information from 47 areas and remotely-detected land surface temperature, precipitation, and phantom records as indicators. The best model recognized six clusters, and the areas in each bunch had unmistakable reactions to the ecological indicators. They presume that the spatial delineation can work on the attack of earth driven illness models, and hereditary calculations give a reasonable and successful methodology for recognizing these clusters.

* 1. **PROPOSED SYSTEM**

This project aims to design time series based ARIMA model using diffferent parameters such as temperature, vegetation and rainfall. Then dimension reduction is done for better prediction of dengue outbreak.

Comprehension of the connection among environment and dengue elements can further develop research drives and asset allotment to assist with battling hazardous pandemics.

A dataset containing environmental factors and weekly dengue cases of san juan city. After performing the data set cleaning and performing feature selection, we have determined that we will be using (min, max, avg) temperature, total precipitation, average and, maximum atmosphere temperature, arithmetic-mean and maximum humidness. These environmental factors for prediction. From the graph obtained by time-series forecast, we interpolate the number of predicted cases and compose with actual cases per week to determine the model efficiency.

**CHAPTER 3**

**DESIGN**

**3.1 SYSTEM REQUIREMENT AND SPECIFICATIONS**

The System Requirement Specifications (SRS) file specifies each and every requirement for the software programme product, as well as the external interfaces to hardware and firmware.

**3.1.1 SOFTWARE REQUIREMENTAND SPECIFICATIONS**

The list of software required for demonstrating our project is:

* 0perating System (OS): Windows 7 & above
* Python
* Python packages: Numpy, Scikitlearn, Pandas, matplotlib, keras, tensorflow, seaborn.

**3.1.2 HARDWARE REQUIREMENT SPECIFICATIONS**

The hardware requirements for the project are:

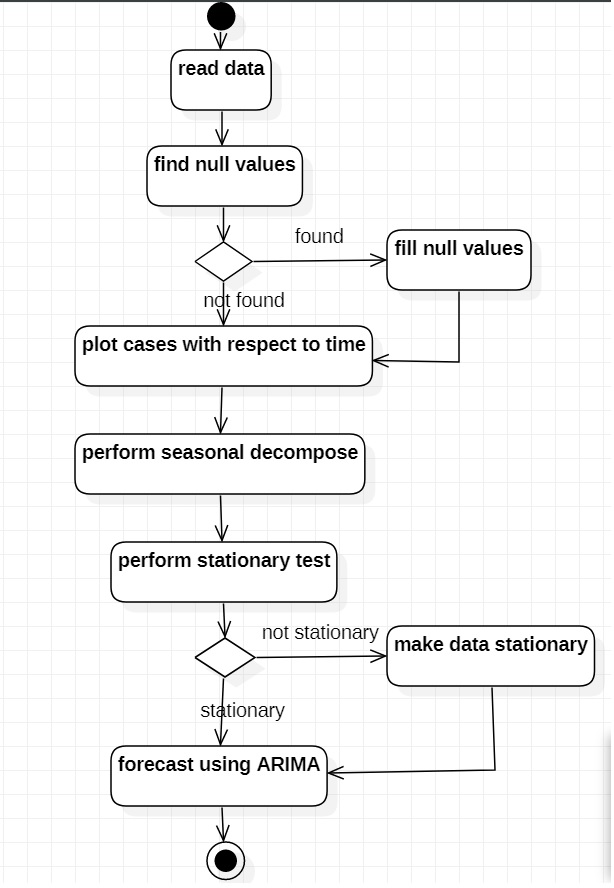
* Processor- Dual Core
* Hard Disk- 50 GB
* RAM- 1 GB

**3.2 TIME SERIES MODELING**

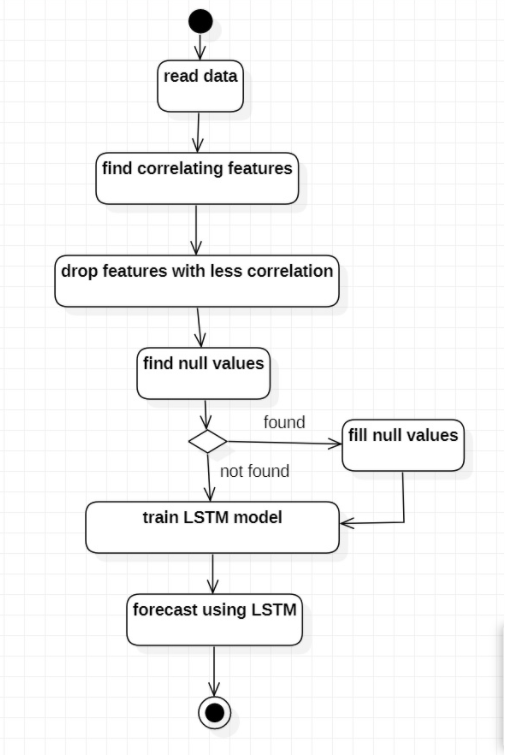
In reality, a time-collection version can be defined as "a set of observations x, each of which is recorded at a predetermined time t." Separate-time series, in which compliances are made over a fixed period of time, and nonstop-time series, in which compliances are made continuously over a period of time, are the two types of time-collection. The use of time-collection for grievance surveillance or tracking is beneficial in alerting the public health government of an impending criticism outbreak and providing ample time for grievance outbreak mitigation sweats to be implemented. A study of emergency-branch literal records found that the use of time-collection methods could be used to detect out-of-the-ordinary patterns, and this work suggested that it could be used as an early-warning device for criticism outbreak. The exigency- branch literal data passed high- delicacy for respiration-affiliated and common pediatric exigency cases read, and turned into touted to be an effective early-- caution device for acts of bioterrorism. Time-series has been used efficaciously for the vaticination of epidemic of specific contagious conditions.

**3.3 UML DIAGRAM**

**3.3.1 Activity Diagram**

****

**Fig 3.1: Activity diagram for ARIMA Model**

****

**Fig 3.2: Activity diagram for LSTM Model**

**CHAPTER 4**

**MODULES**

* 1. **IMPORTIG DATA**
* **pandas.read\_csv(filepath\_or\_buffer)**

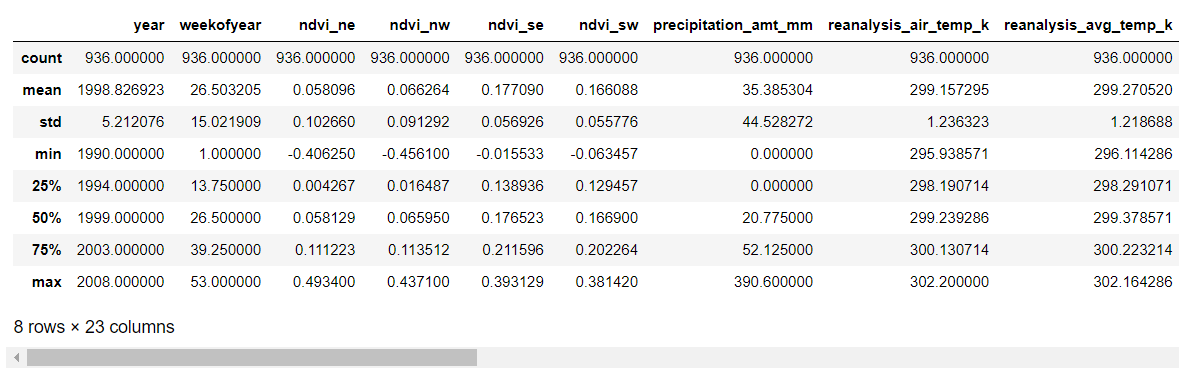
**Record-direction or buffer:str, course item or document-like object** Any sane string direction is acceptable. It's possible that the string is a URL. http, ftp, s3, gs, and file are all valid URL schemes. A host is required for record URLs.****

**Fig 4 .1: Importing data**

Here, we have

In the dataset we have collected, it has the data for two cities San Juan and Iquitos. So, we have split the dataset and took only the San Juan city.

* + 1. **Describing the Data**

****

**Fig 4.2: Describing data**

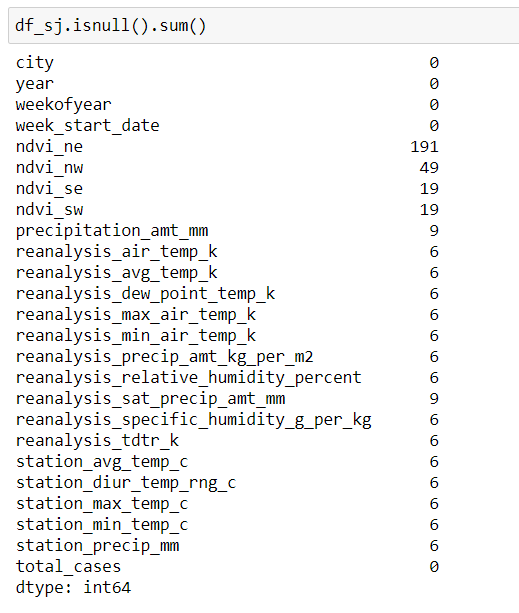
In the above dataset we have data from the year 1900 to the year 2008.

* 1. **DATA PRE - PROCESSING**

**4.2.1 Checking null values**

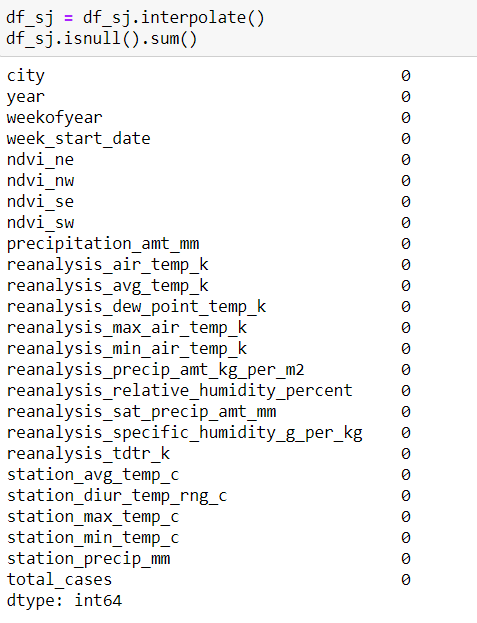
Data Cleaning is one of the important steps. Data cleaning can be done in many ways. One of them is handling missing values.Here in the dataset first we have to identify null values and then we have to fill or replace them.

**df.isnull().sum()** will return the count of null values in each column.



**Fig 4.3: Checking Null values**

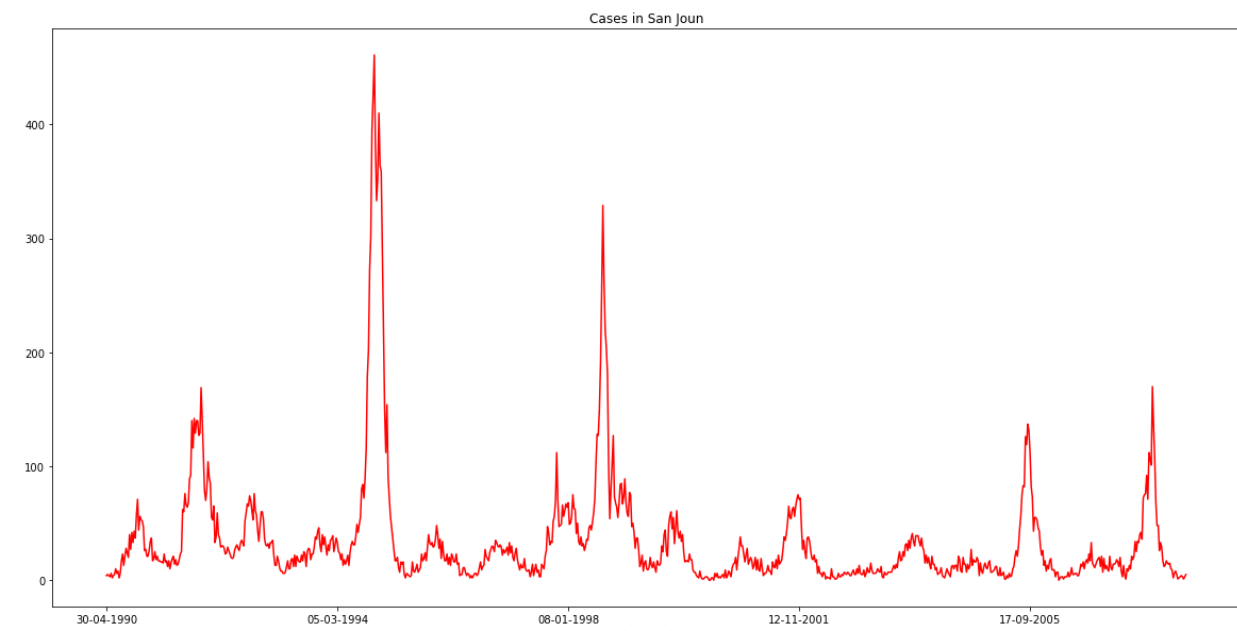
To fill the missing values here we have used interpolate() method.



**Fig 4.4: Filling Null values**

* 1. **EXPLORATORY DATA ANALYSIS**

This the graph we obtain when we plot by considering the number of cases per week w.r.t the week start date.

****

**Fig 4.5: Cases in San Juan**

**4.3.1 Seasonal Decompose**

It is enjoyable to break down a time series into its systematic and non-systematic additives in order to select a forecasting model.

Systematic: Time collection components with consistency or recurrence that can be described and modelled.

Non-Systematic: Time collection components that cannot be directly modelled.

Decomposition is the name for this technique.

It provides a useful summary model that can be used to think about a time collection forecasting problem in a structured way, both in terms of modelling complexity and specifically in terms of how to first-class capture those additives in a given version.

A time series is thought to contain two systematic components, fashion and seasonality, as well as one non-systematic issue, noise.

Within the collection, there is a trend of increasing or decreasing fees.

Seasonality is the collection's recurring short-term cycle.

The random variation within the series is referred to as noise.

These elements must be considered and addressed at some point during facts practise, model selection, and model tuning.

We used the statsmodels library, which has a feature called seasonal decompose that implements the naive, or classical, decomposition method ().

In time series analysis, this method is used to represent a time series as a sum or made up of three additives: linear fashion, periodic (seasonal) element, and random residuals.

This property necessitates specifying whether the model is additive or multiplicative

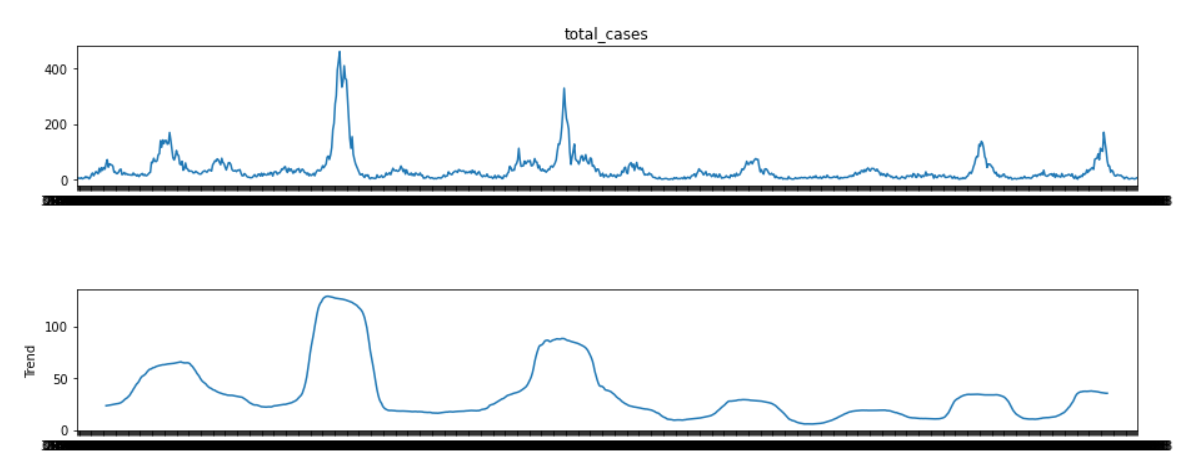
In an additive model, components are introduced in the following order:

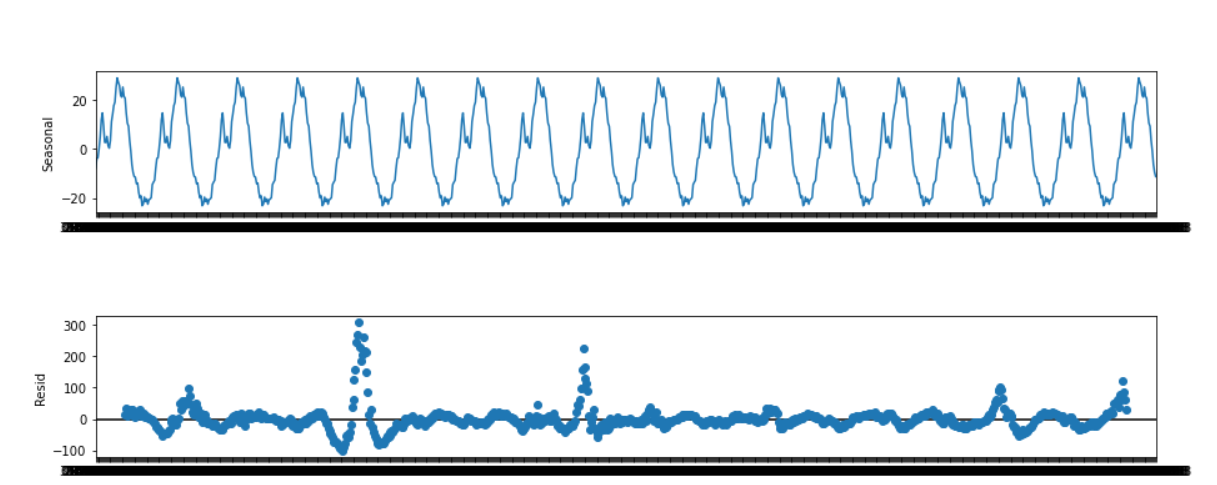
Level + Trend + Seasonality + Noise = y(t).

According to a multiplicative model, the components are collectively extended as follows:

Noise \* Level \* Trend \* Seasonality = y(t).

A result item is returned by the seasonal decompose() function. The final result object is made up of arrays that allow you to access four different statistics from the decomposition. Which can be graphed and analysed?





**Fig 4.6: Seasonal Decomposition Graph**

**4.4STATIONARY TEST**

The main reason for checking stationarity for a that stationary processes are easier to analyse.

If the arithmetic mean and standard deviation of a time series dataset doesn’t change over a period of time i.e. the way a series changes does not change over time by this we mean rate of change is same over a period of time.

In order to identify stationarity, we have many methods but in this module we will be using two methods:

The first being visualisation, in this method we plot the original time series data with its rolling mean and rolling standard deviation. based on the plot obtained we compare and identify whether it is stationary or not.

The other method is more of a statistical test called dickey-fuller test, also known as unit root test.

This test basically tells us how strongly a time series is defined by a trend.

We decipher this outcome utilizing the term p from the test. A p-value under an edge, (for example, 5% or 1%) proposes we reject the invalid theory (fixed), in any case, a p-value over the limit recommends we neglect to dismiss the invalid speculation (non-fixed).

p-value > 0.05: Fail to dismiss the invalid theory (H0), the information has a unit root and is non-fixed.

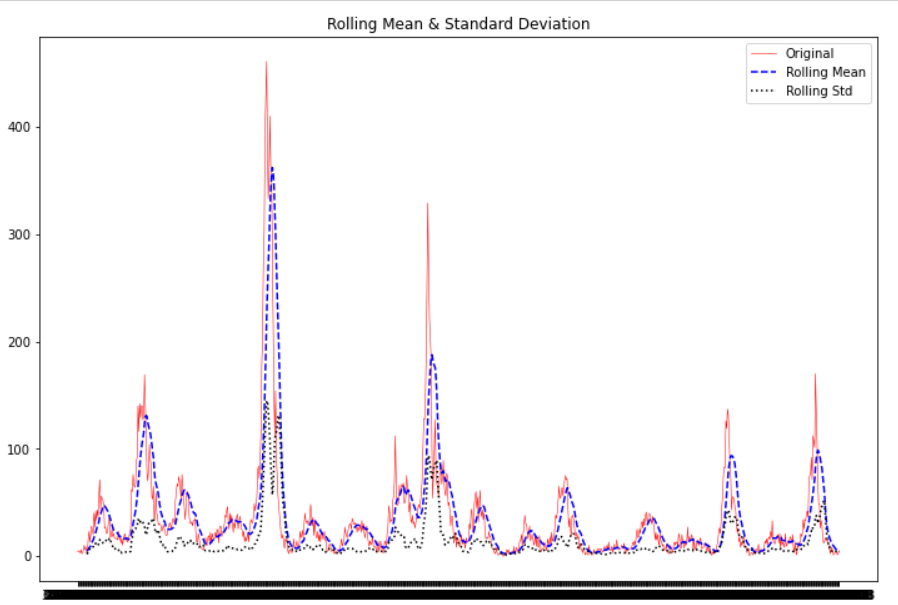
p-value <= 0.05: Reject the invalid theory (H0), the information doesn't have a unit root and is fixed.

the more negative the test statistic the more chance of it being stationary.

**4.4.1 VISUALIZATION**



**Stationary test by visualization:**

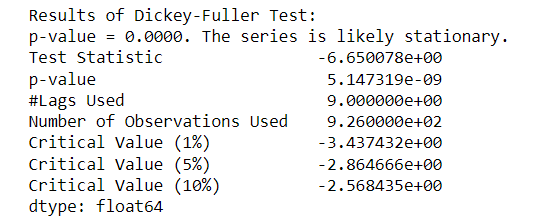


**Fig 5.1: Rolling Mean & Standard Deviation**

Here we have plotted the original graph (number of cases w.r.t time) rolling mean graph and rolling standard deviation graph.

By comparing the three graphs we can see that the graph is mostly stationary

**4.4.2 DICKEY FULLER TEST**



**Fig 5.2: Dickey-Fuller Test Results**

Dickey-Fuller Test is also known as unit root test.

A unit root test tests whether a period series isn't fixed and comprises of a unit root in time series investigation. The presence of a unit root in time series characterizes the invalid theory, and the elective speculation characterizes time series as fixed.

This test basically tells us how strongly a time series is defined by a trend.

We decipher this outcome utilizing the term p from the test. A p-value under an edge, (for example, 5% or 1%) proposes we reject the invalid theory (fixed), in any case, a p-value over the limit recommends we neglect to dismiss the invalid speculation (non-fixed).

p-value > 0.05: Fail to dismiss the invalid theory (H0), the information has a unit root and is non-fixed.

p-value <= 0.05: Reject the invalid theory (H0), the information doesn't have a unit root and is fixed.

The more negative the test statistic the more chance of it being stationary.

**4.5ARIMA IMPLEMENTATION**

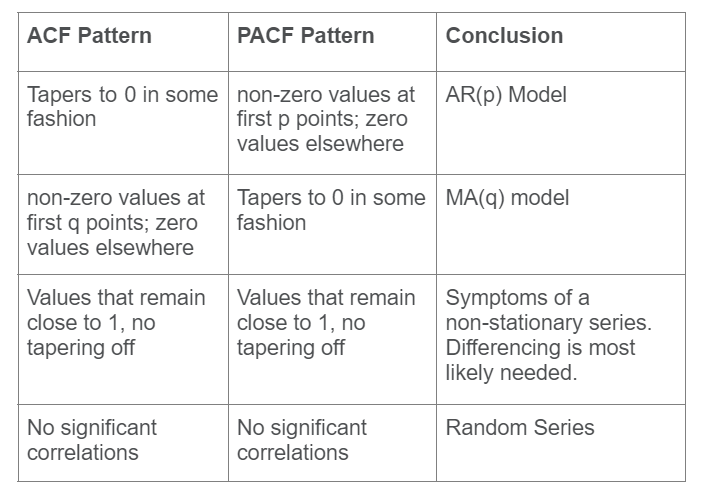
**Autoregressive Integrated Moving Average (ARIMA)**

ARIMA, or autoregressive integrated moving average, is a statistical analysis model that uses time series data to better understand the data set or anticipate future trends.

Autoregressive statistical models anticipate future values based on past values. For example, an ARIMA model might attempt to estimate a company's earnings based on past performance or predict a stock's future price based on past performance.

An autoregressive integrated moving average model is a type of regression analysis that assesses the importance of one dependent variable in relation to other changing variables. The model's purpose is to forecast future securities or financial market movements by analysing discrepancies between values in the series rather than actual values.

Each of the components of an ARIMA model can be understood by outlining them as follows:

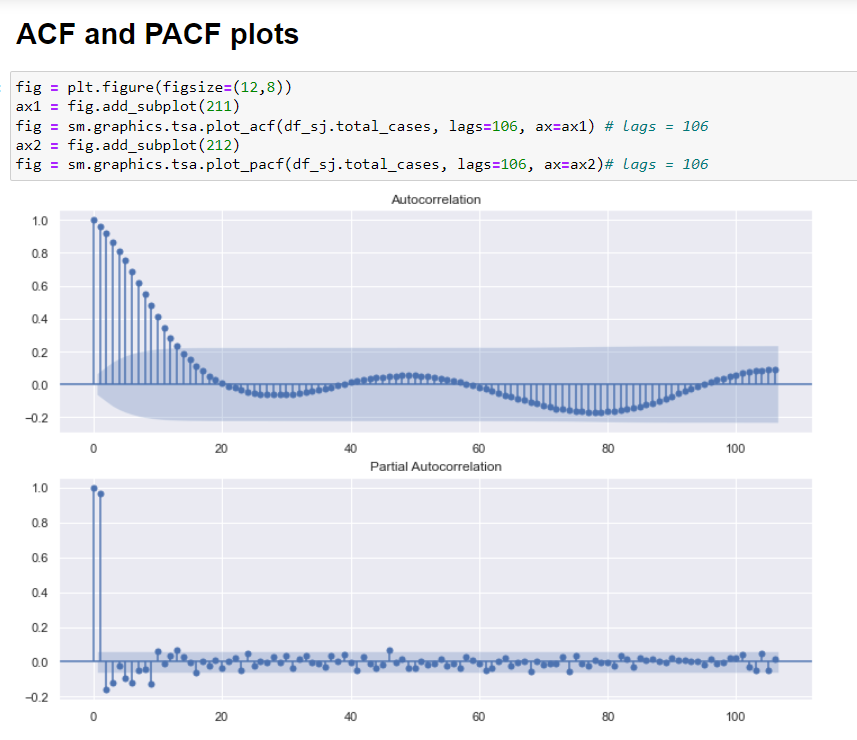
* Autoregression (AR): a model in which a changing variable regresses on its own prior, or previous values.
* Integrated (I): the difference between raw observations and previous values that allows the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).
* A moving average (MA) takes into account the relationship between an observation and a residual error from a moving average model applied to lagged observations.
* 

To be stationary, the data in an autoregressive integrated moving average model must be different. A stationarity model is one that demonstrates that the data is consistent throughout time. Because most economic and market data exhibit trends, the goal of separating is to eliminate any such trends or seasonal structures.

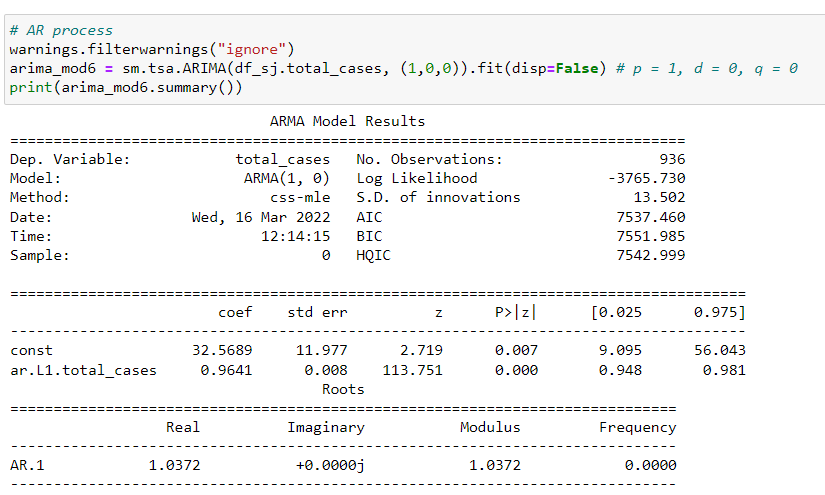
Seasonality, or when data exhibits regular and predictable patterns that reoccur over the course of a calendar year, may have a detrimental impact on the regression model. Many of the computations during the process cannot be made efficiently if a trend arises and stationarity is not visible.

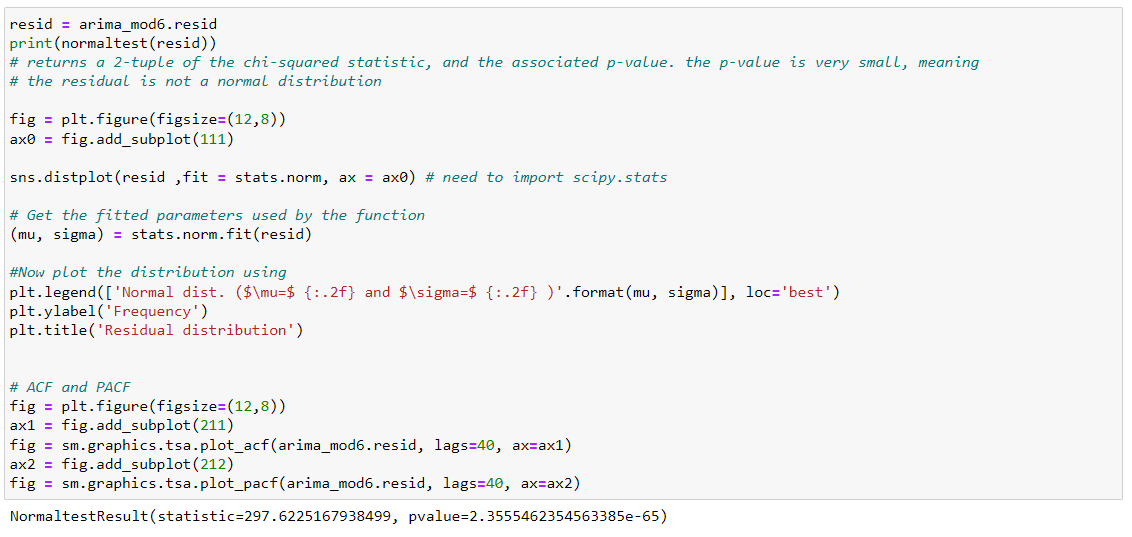
**4.5.1 ACF & PACF PLOTS**

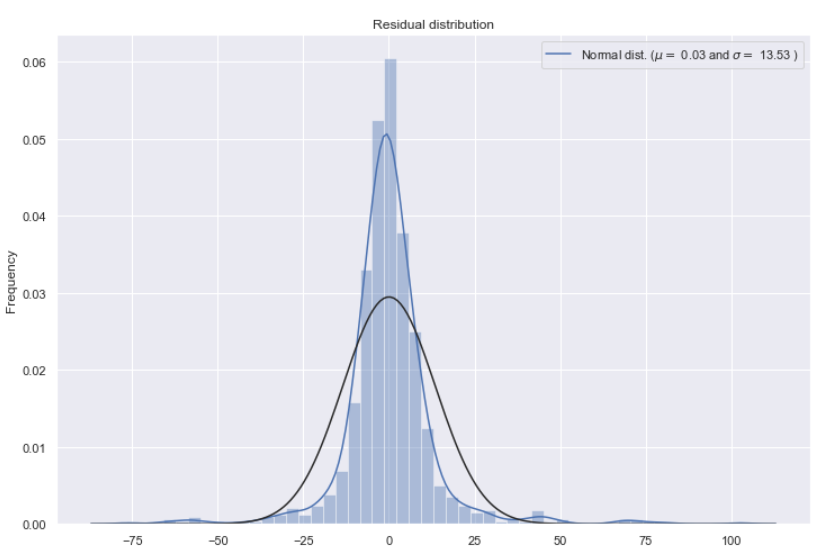
A time series is created when a single variable is measured repeatedly across time. In many different ways, time series values can be associated with earlier values in the same series. A time series' temporal dynamics are represented by such a correlation across time, which is commonly referred to as autocorrelation. Researchers frequently employ a statistical model class known as "autoregressive integrated moving." The average (ARIMA) model is used to investigate the temporal dynamics of a single time series. They must first investigate the ACF and PACF for the series. An ACF and a PACF may also aid researchers in determining whether there is any over-time correlation in the residuals of a typical time series study.

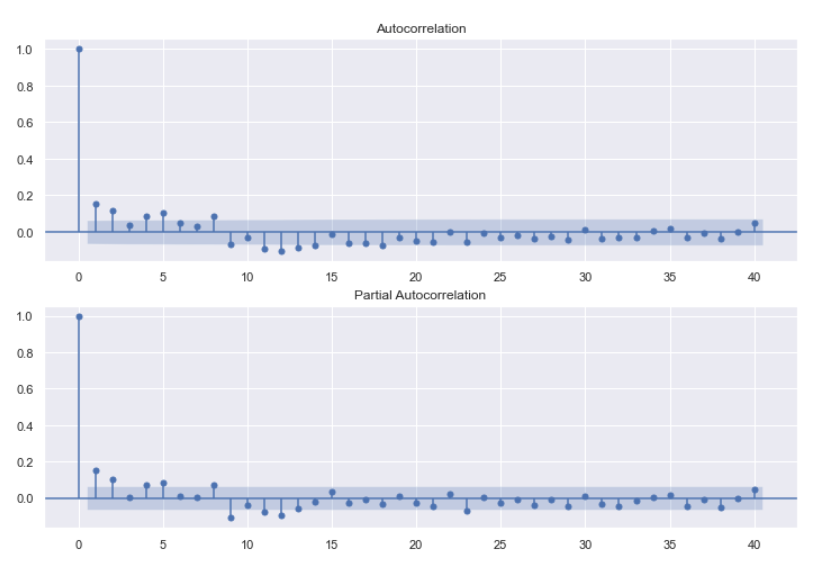


**4.5.2 MODEL BUILDING**

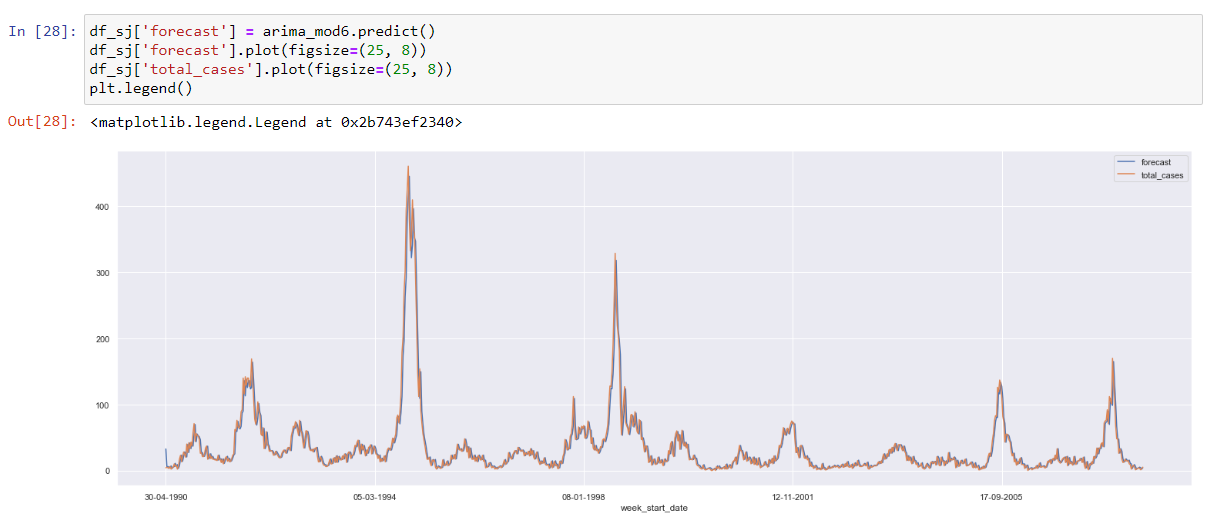
****

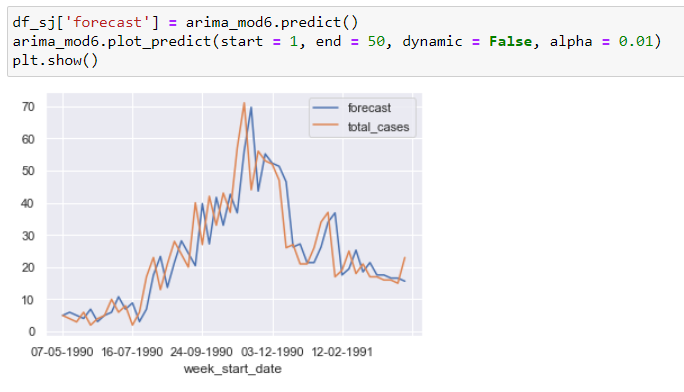
****

****

****

**PREDICTION**





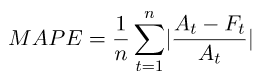
**4.5.3 MODEL EVALUATION**

we can calculate the average error obtained by the forecast using the Mean Absolute Error formula:



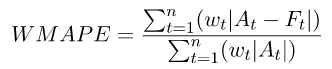
**MAPE**

MAPE is perhaps the most well-known strategy to quantify gauge exactness. It implies Mean Absolute Percentage Error and it estimates the rate blunder of the conjecture according to the genuine qualities. As it ascertains the typical mistake after some time or various items, it doesn't separate between them. This implies that it accepts no inclination between what day for sure item to foresee better. It is determined as follows:



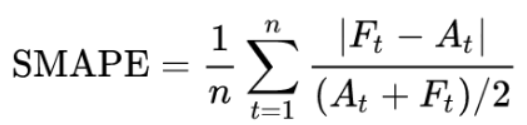
**WMAPE**

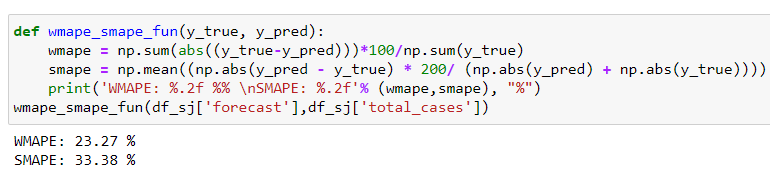
It means Weighted Mean Absolute Percentage Error and is calculated as follows:



**SMAPE**

It stands for Symmetric Mean Absolute Percentage Error and is calculated as:





**LSTM (LONG SHORT-TERM MEMORY MODEL) IMPLEMENTATON**

LSTM is a class of recurrent neural network.

**Neural Networks:**

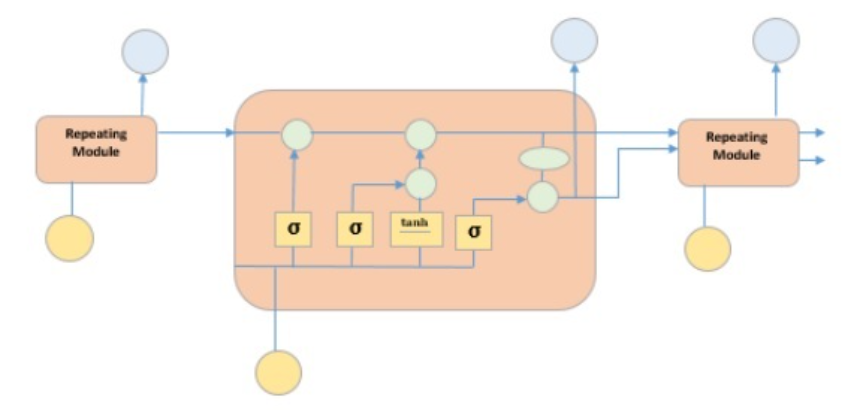
An artificial neural network is a layered structure of connected neurons, inspired by biological neural networks. It is not one algorithm but combinations of various algorithms which allows us to do complex operations on data.

**Recurrent Neural Networks:**

It is a class of brain networks customized to manage fleeting information. The neurons of RNN have a cell state/memory, and info is handled by this inward state, which is accomplished with the assistance of circles with in the brain organization. There are repeating module(s) of 'tanh' layers in RNNs that permit them to hold data. In any case, not for quite a while, which is the reason we really want LSTM models.

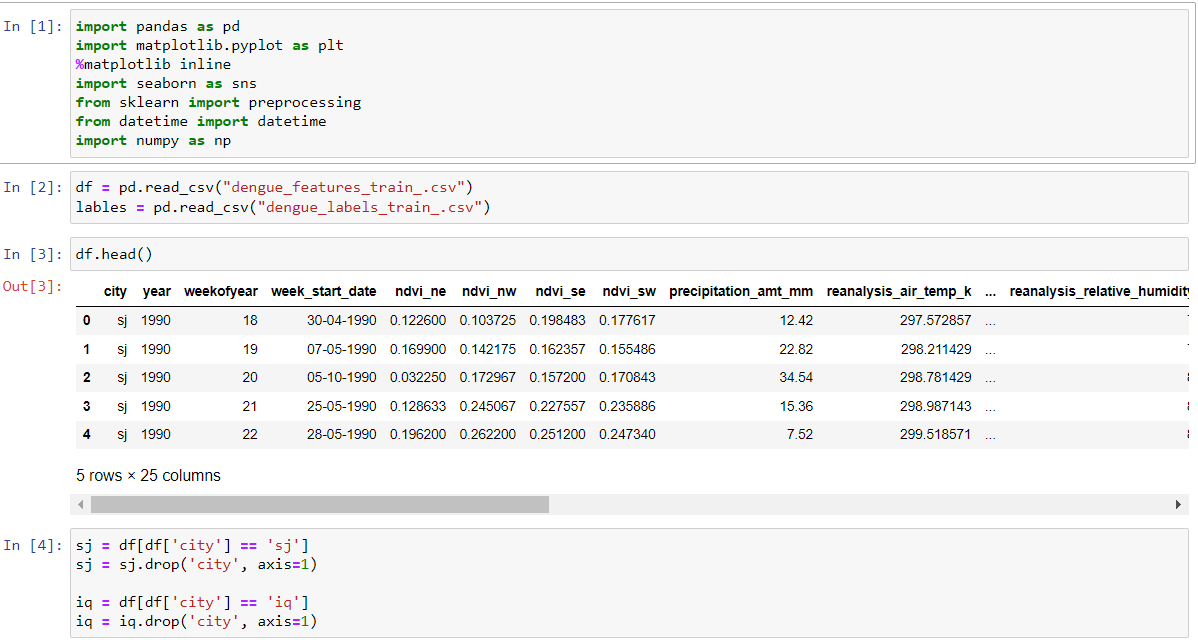
**LSTM:**

It is special kind of recurrent neural network that is capable of learning long term dependencies in data. This is achieved because the recurring module of the model has a combination of four layers interacting with each other.



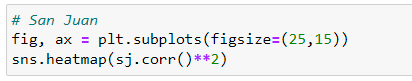
The image above portrays four brain network layers in yellow boxes, point wise administrators in green circles, contribution to yellow circles and cell state in blue circles. A LSTM module has a cell state and three doors which furnishes them with the ability to learn, forget or hold data from every one of the units specifically. The cell state in LSTM assists the data with coursing through the units without being modified by permitting a couple of direct communications. Every unit has an info, yield and a neglect entryway which can add or eliminate the data to the cell state. The neglect entryway concludes which data from the past cell state ought to be forgotten for which it utilizes a sigmoid capacity. The info entryway controls the data stream to the ongoing cell state utilizing a point-wise duplication activity of 'sigmoid' and 'tanh' individually. At long last, the result door concludes which data ought to be given to the following secret state.

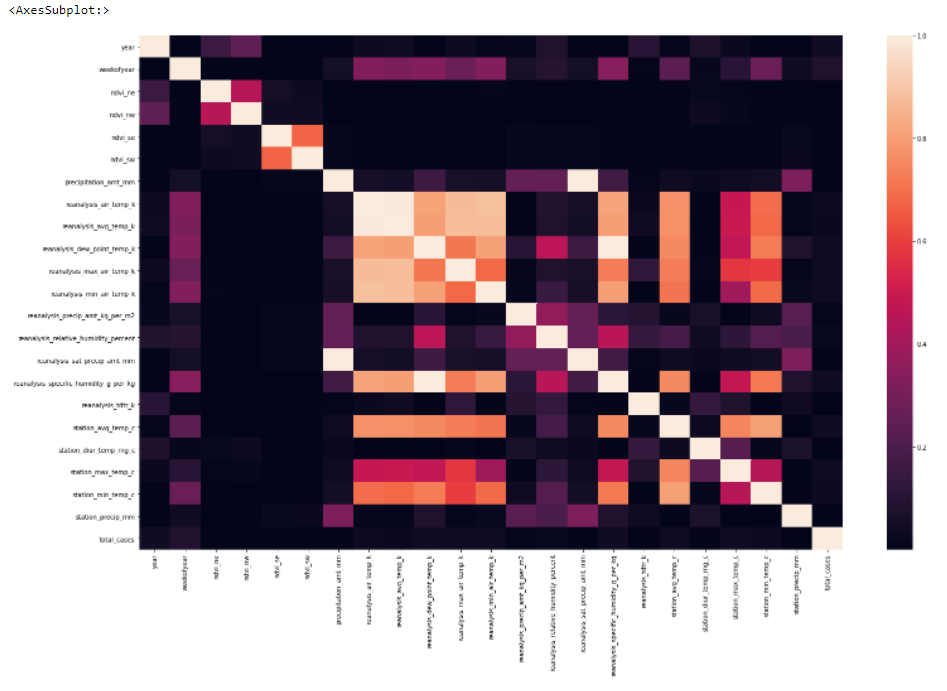
**4.6.1 IMPORTING DATA**

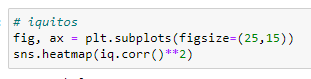
****

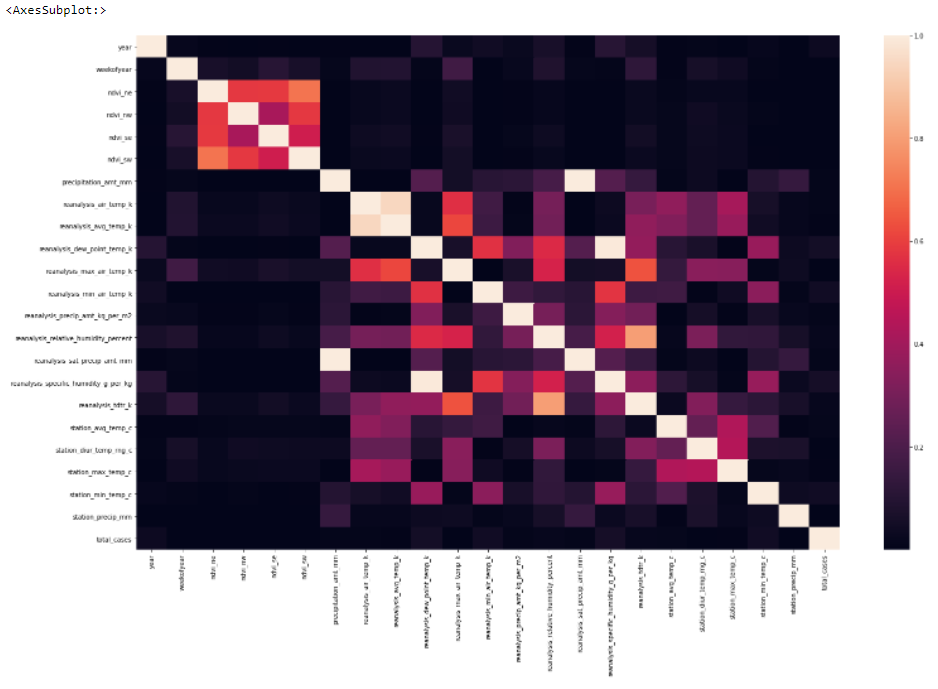
**4.6.2 HEAT MAP**

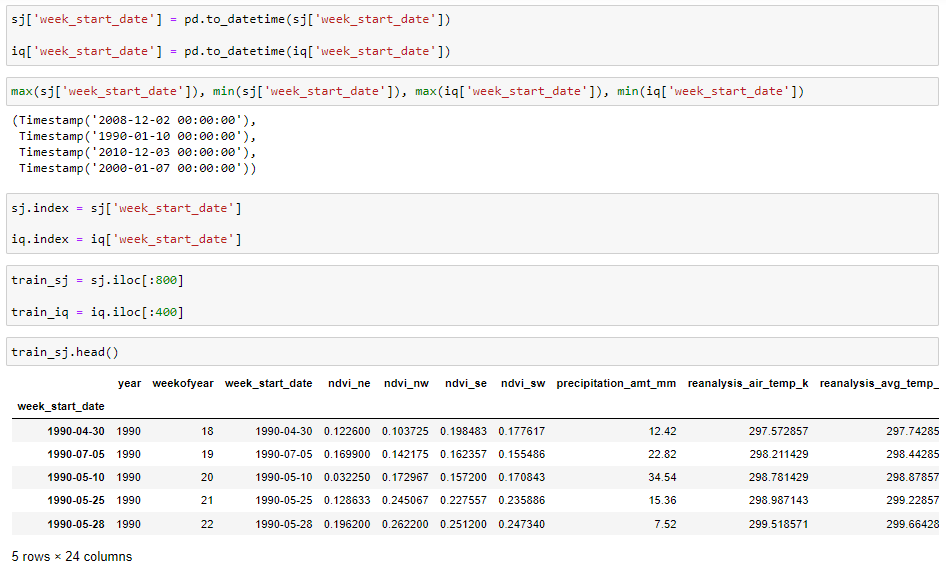
A heat map (or heatmap) is an information representation strategy that shows size of a peculiarity as variety in two aspects. The variety in variety might be by shade or power, giving clear obvious prompts to the per user about how the peculiarity is bunched or changes over space. There are two essentially various classifications of intensity maps: the bunch heat map and the spatial intensity map**.**

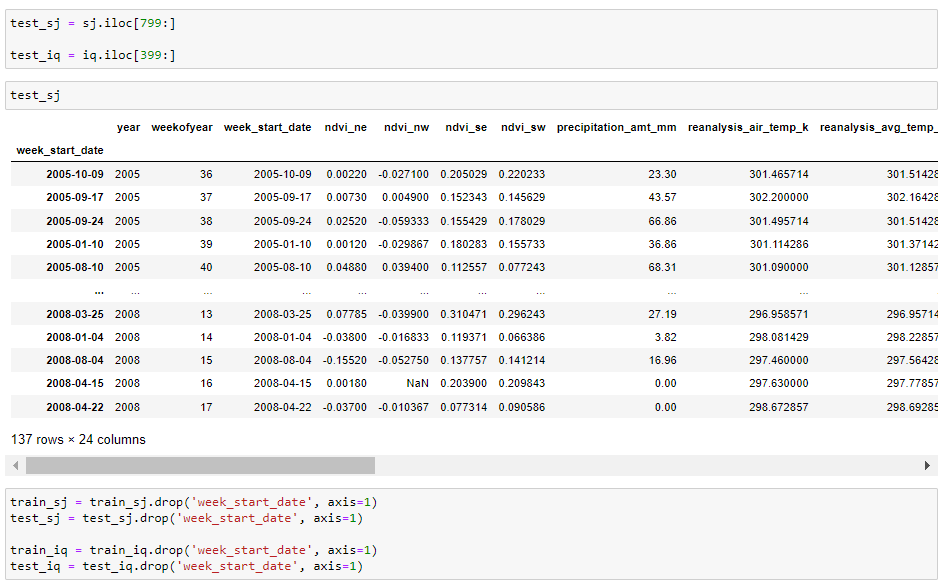
****

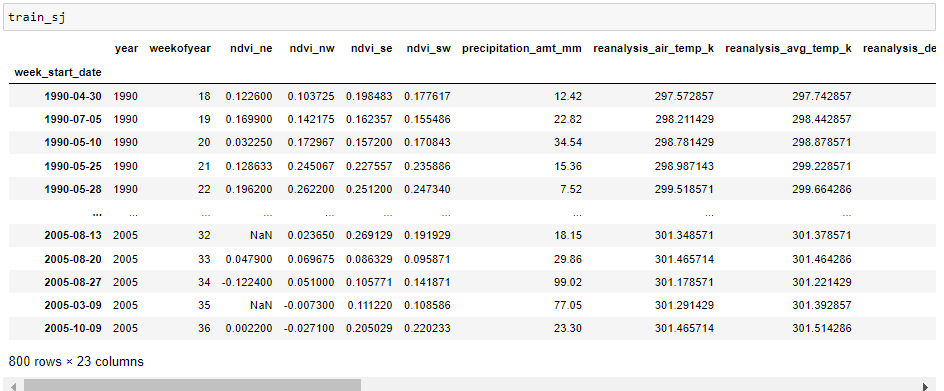
****

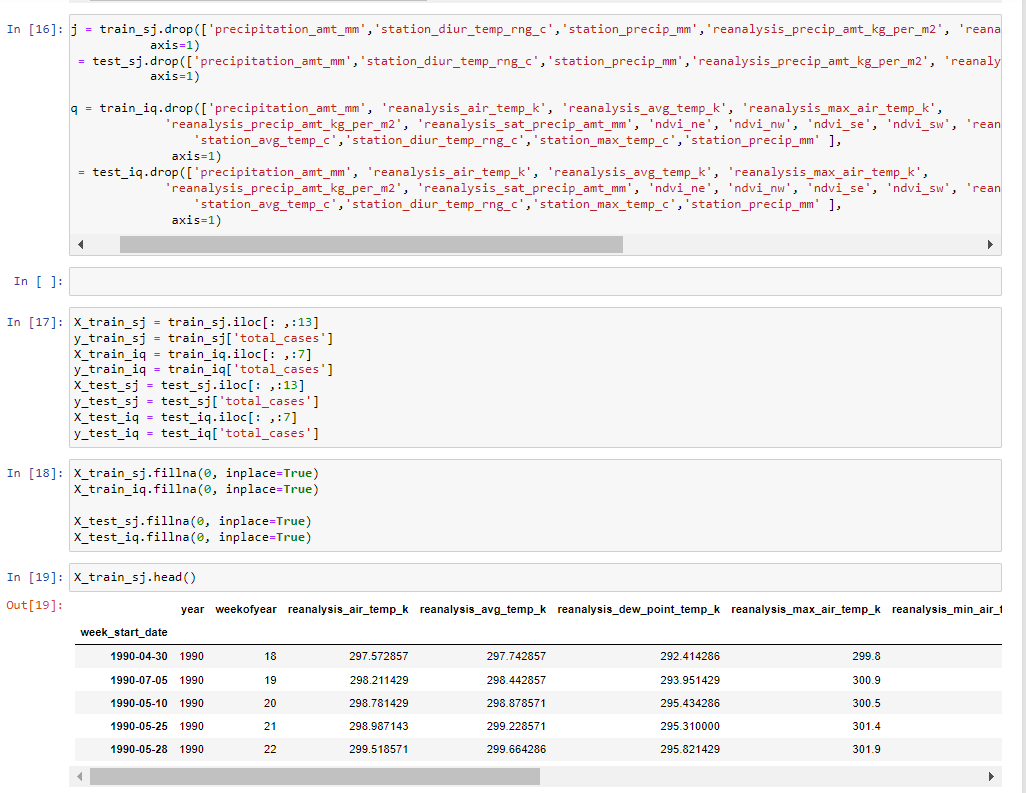
****

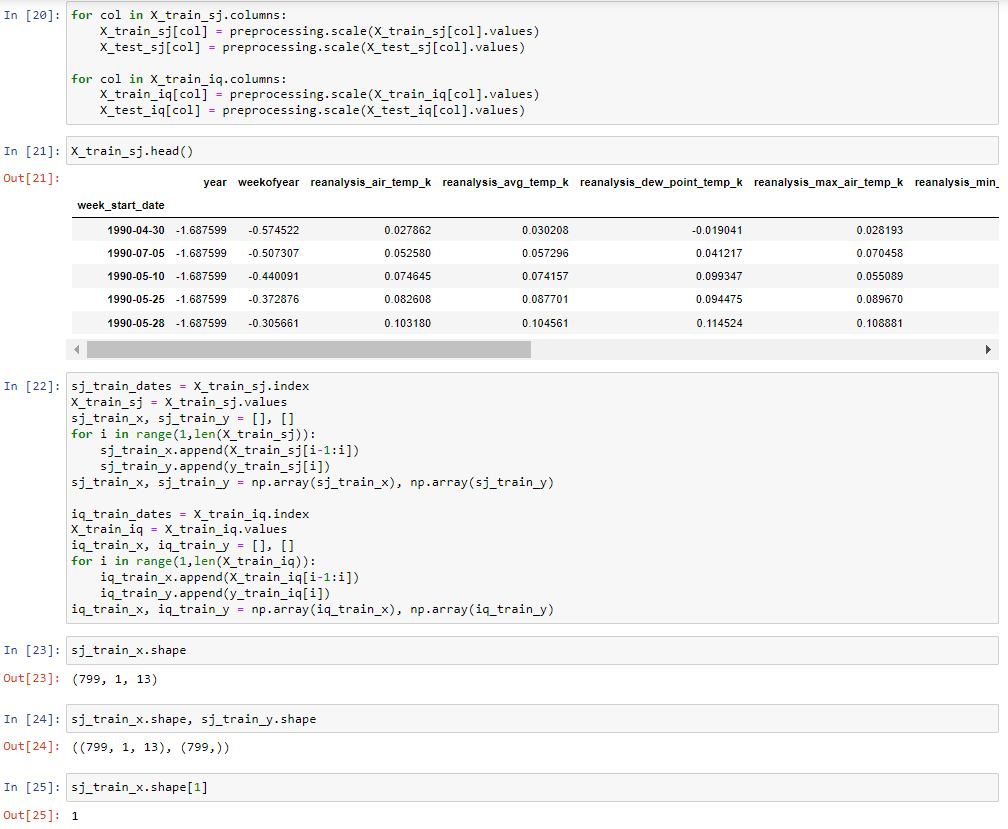
****

****

****

****

****

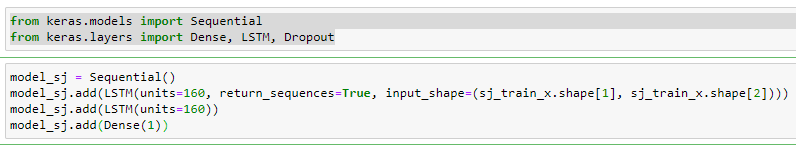
****

**4.6.3 FORECAST FOR SAN JAUN**

TensorFlow is a start to finish open-source stage for AI. It has a thorough, adaptable biological system of instruments, libraries and local area assets that allows specialists to push the cutting edge in ML and designers effectively construct and send ML fueled applications.

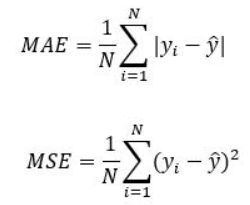
Keras is a deep learning API written in Python, running on top of the AI stage TensorFlow. It was created with an attention on empowering quick trial and error. Having the option to go from thought to result however quick as conceivable seems to be vital to doing great exploration.

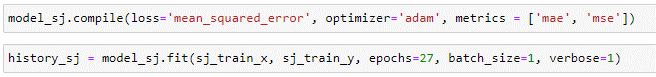
The sequential model is the linear stack of layers. You can create a Sequential model by passing a list of layer instances to the constructor or You can also simply add layers via the .add() method.

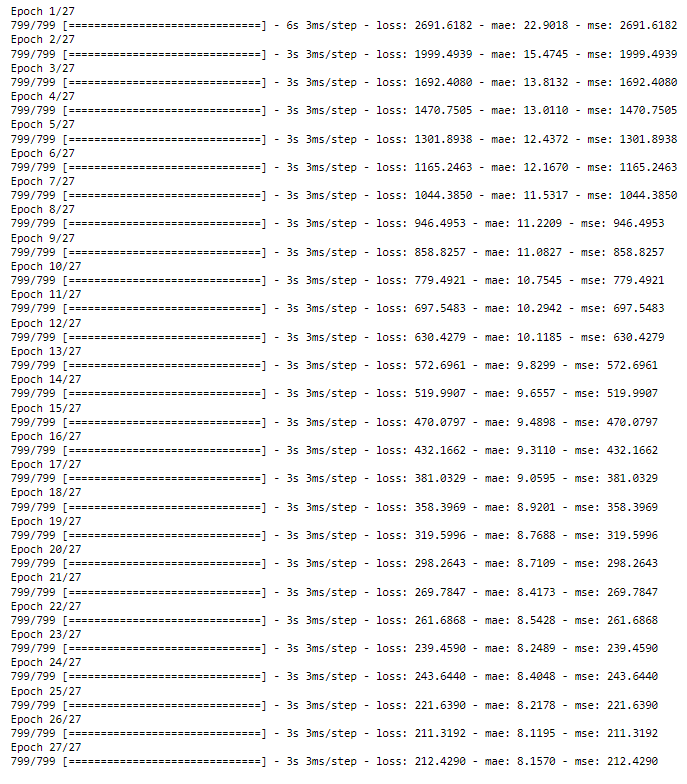


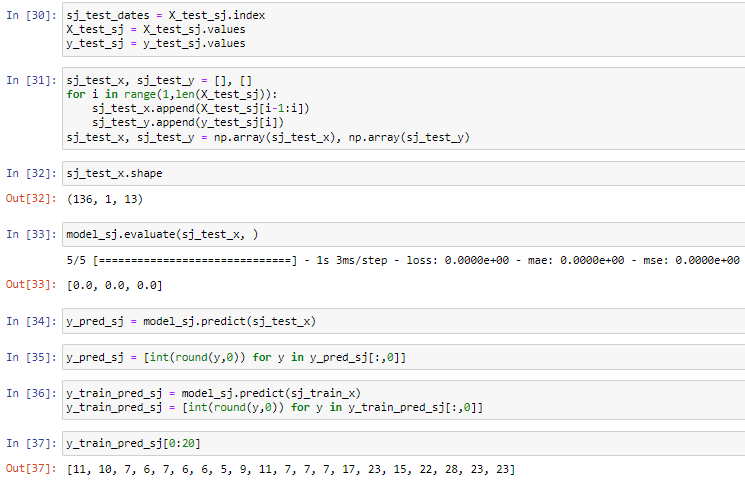
MAE (Mean absolute error) represents the difference between the original and predicted values extracted by averaged the absolute difference over the data set.

MSE (Mean Squared Error) represents the difference between the original and predicted values extracted by squared the average difference over the data set.

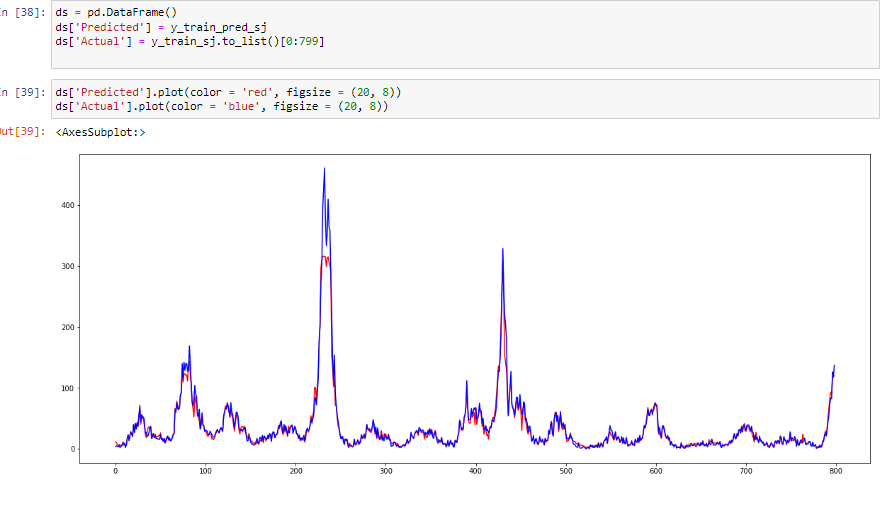




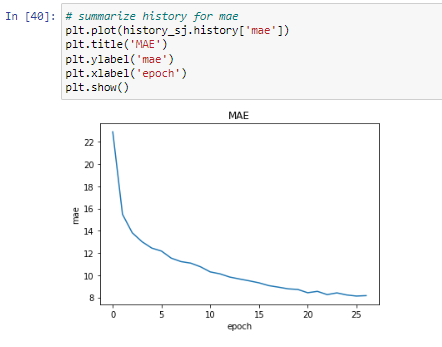


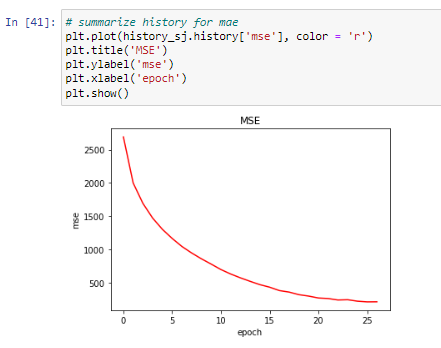


**PLOTTING FOR SANJAUN DATA**

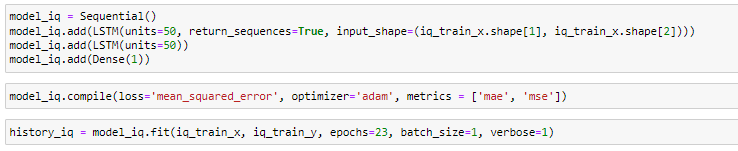
****

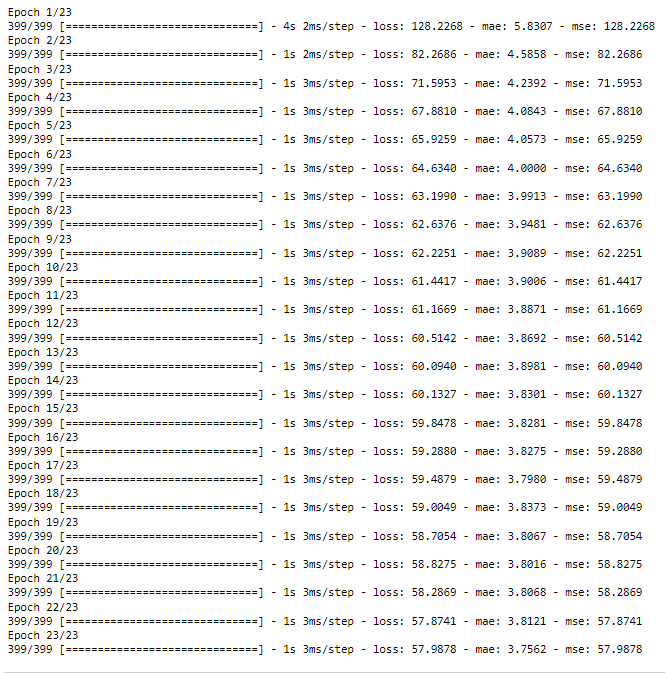
**SANJAUN MODEL EVALUATION**

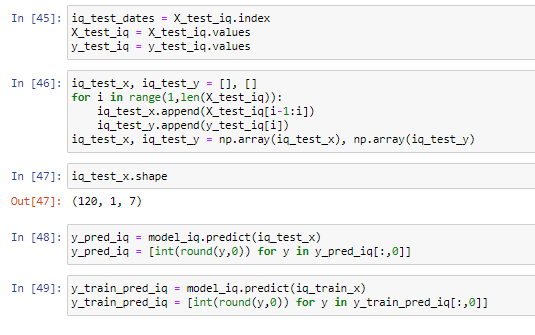




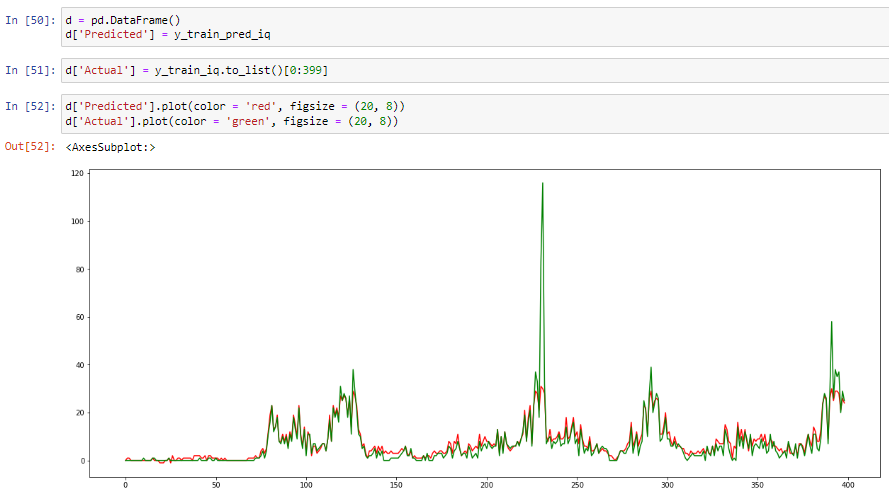
**4.6.4 FORECAST FOR IQUITOS DATA**

****

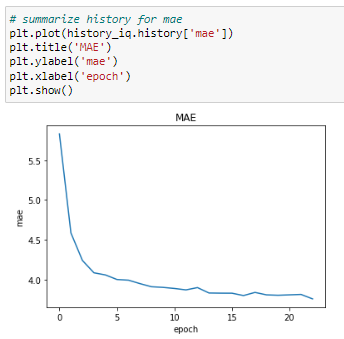
****

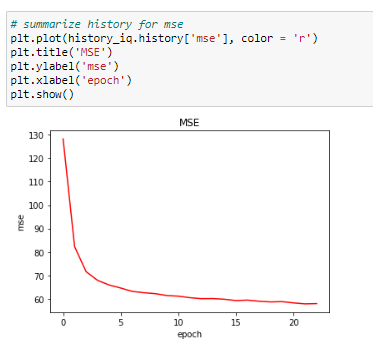
****

**PLOTTING FOR IQUITOS DATA**

****

**IQUITOS MODEL EVALUATION**

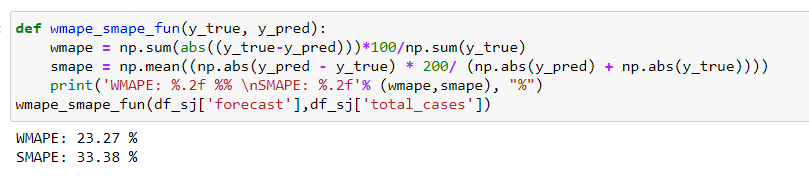
****

****

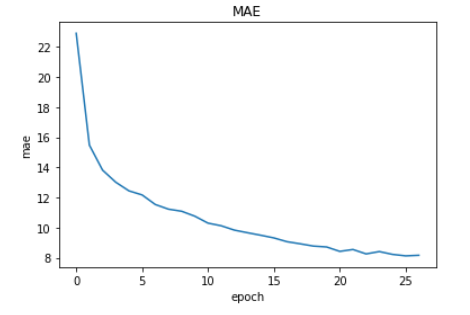
**CHAPTER 5**

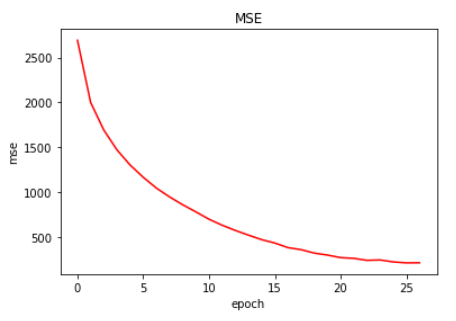
**RESULTS**

**ARIMA MODEL**

****

**LSTM MODEL EVALUATION**





**CHAPTER 6**

**CONCLUSION AND FUTURE SCOPE**

Our analyzation of the information has revealed that the total collected Dengue Fever cases in San Juan exhibit/has shown us a trend and seasonal sample, which is the best-in-class forecast using an ARIMA model. The work done here validates the effort to make public-interest statistics open, as it encourages combined work to discover new knowledge that can benefit public policy and decision-makers. We specifically hope that public health officers will be encouraged to share their data as a way to tap into the expertise of epidemiology researchers.

In regards to the future work, we intend to investigate the impact of out-of-pattern statistics on models, as well as conduct correlation studies using meteorological data.

This study applied a profound learning way to deal with anticipating the following month to month all out of revealed dengue cases in san juan City. The LSTM model has exhibited its capacity to hold data from a long grouping of information. The model created eminently more exact conjectures than the different methodologies by proficiently alluding to data from the beyond 12 mo to draw its expectation. Of the model ascribes considered, the time step boundary shown the most impact on the figure exactness. This hyperparameter directs the contribution to the model, consequently affirming the autoregressive quality of dengue case counts. Thinking about more than one secret layer is additionally superfluous to create exact figures. The review showed a prominent expansion in estimate exactness utilizing the profound learning approach; in any case, there is still a lot of opportunity to get better. This review utilized a generally little dataset contrasted with the average fruitful utilizations of AI calculations, so it is recommended to utilize a bigger dataset when accessible. The LSTM model design could likewise be further inspected by tweaking other model hyperparameters such as expanding the quantity of ages, changing the learning rate, and thinking about other analyzers. The utilization of other succession model organizations, for example, the gated intermittent units, which is another RNN network expansion, or the transformer models could likewise be investigated. Expanding on a streamlined model, the utilization of move learning can likewise be researched to work on the estimate exactness for different districts with less dengue cases.

The secret layer LSTM and the enactment work ReLU are applied in the created model. The RMSE of other machine it is additionally figured to learn calculations furthermore, it is clear that the LSTM displays better execution. Despite the fact that the expectation level is not over 95%, it is apparent that the likeliness of the contamination of dengue cases is expanding step by step and it can be utilized as an indication of caution to the Government to take proper activities/ventures towards the avoidance of the infection. The prepared model can be stretched out to move learning. The precision level of the anticipated model can be improved by considering the extra datasets for over a time of 25 years. Further the model can be tried by expanding the number of ages, other initiation capacities. The model is executed in a GPU climate, the presentation can be tried by executing the model in TPU equipment.

**REFERENCES**

1. Choudhury, Z.M., Banu, S., Islam, A.M.: Forecasting dengue incidence in Dhaka, Bangladesh: A time series analysis. (2008).
2. Lal, V., Gupta, S., Gupta, O., Bhatnagar, S.: Forecasting incidence of dengue in Rajasthan, using time series analyses. Indian J. Public Health. 56, 281 (2012).
3. Earnest, A., Tan, S.B., Wilder-Smith, A., Machin, D.: Comparing Statistical Models to Predict Dengue Fever Notifications. Comput. Math. Methods Med. 2012, e758674 (2012)
4. Chadsuthi, S., Modchang, C., Lenbury, Y., Iamsirithaworn, S., Triampo, W.: Modelingseasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using time–series and ARIMAX analyses. Asian Pac. J. Trop. Med. 5, 539–546(2012).
5. Luz, P.M., Mendes, B.V.M., Codeço, C.T., Struchiner, C.J., Galvani, A.P.: Time Series Analysis of Dengue Incidence in Rio de Janeiro, Brazil. Am. J. Trop. Med. Hyg. 79, 933–939 (2008).